Relative complexity in scientific discourse

Hundt, Marianne ; Denison, David ; Schneider, Gerold

Abstract: Variation and change in relativization strategies are well documented. Previous studies have looked at issues such as (a) relativizer choice with respect to the semantics of the antecedent and type of relative, (b) prescriptive traditions, (c) variation across text types and regional varieties, and (d) the role that relative clauses play in the organization of information within the noun phrase. In this article, our focus is on scientific writing in British and American English. The addition of American scientific texts to the ARCHER corpus gives us the opportunity to compare scientific discourse in the two national varieties of English over the whole Late Modern period. Furthermore, ARCHER has been parsed, and this kind of syntactic annotation facilitates the retrieval of information that was previously difficult to obtain. We take advantage of new data and annotation to investigate two largely unrelated topics: relativizer choice and textual organization within the NP. First, parsing facilitates easy retrieval of relative clauseswhichwere previously difficult to retrieve from plain-text corpora by automatic means, namely that- and zero relatives. We study the diachronic change in relativizer choice in British and American scientific writing over the last three hundred years; we also test for the accuracy of the automatically retrieved data. In addition, we trace the development of the prescriptive aversion to which in restrictive relatives (largely peculiar to American English). Second, the parsed data allow us to investigate development in the structure of the NP in this genre, including not only phrasal but also clausal modification of the head noun. We examine the contribution of relative clauses to NP complexity, sentence length and structure. Structural changes within the NP, we argue, are related to the increased professionalization of the scientific publication process.

DOI: https://doi.org/10.1017/S1360674312000032

to be published as

Hundt, Marianne, David Denison & Gerold Schneider. 2012. Relative complexity in scientific discourse. *English Language and Linguistics* 16.2, 209-40.

Relative complexity in scientific discourse

MARIANNE HUNDT

University of Zürich

DAVID DENISON

University of Manchester

GEROLD SCHNEIDER

University of Zürich

Abstract

Variation and change in relativization strategies are well documented. Previous studies have looked at issues such as (a) relativizer choice with respect to the semantics of the antecedent and type of relative, (b) prescriptive traditions, (c) variation across text types and regional varieties, and (d) the role that relative clauses play in the organization of information within the noun phrase.

In this article, our focus is on scientific writing in British and American English. The addition of American scientific texts to the ARCHER corpus gives us the opportunity to compare scientific discourse in the two national varieties of English over the whole Late Modern period. Furthermore, ARCHER has been parsed, and this kind of syntactic annotation facilitates the retrieval of information that was previously difficult to obtain. We take advantage of new data and annotation to investigate two largely unrelated topics: relativizer choice and textual organization within the NP.

First, parsing facilitates easy retrieval of relative clauses which were previously difficult to retrieve from plain-text corpora by automatic means, namely *that*- and zero relatives. We study the diachronic change in relativizer choice in British and American scientific writing over the last three hundred years; we also test for the accuracy of the automatically retrieved data. In addition, we trace the development of the prescriptive aversion to *which* in restrictive relatives (largely peculiar to American English).

Second, the parsed data allow us to investigate development in the structure of the NP in this genre, including not only phrasal but also clausal modification of the head noun. We examine the contribution of relative clauses to NP complexity, sentence length and structure. Structural changes within the NP, we argue, are related to the increased professionalization of the scientific publication process.

1 INTRODUCTION[1]

Relative clauses have attracted scholarly attention regarding their overall structure, different kinds of relativizer, the semantics of the antecedent and the function of the relative clause in relation to it, to name but a few aspects. The focus in the second part of our analyses is on relative clauses as part of noun phrase complexity. We therefore limit our investigation to adnominal relative clauses, i.e. those with an NP as antecedent. Forms that typically relativize an NP and themselves either constitute an NP[2] include *who, whom, whose, which, that* and zero. However, *whose* and *whom* were not included in the parser grammar used to annotate and automatically retrieve the data used for this paper. We therefore concentrate on relative clauses with the relativizers *who, which, that* and zero.

From our study of relative clauses (and some related structures) in scientific discourse we hope to add to knowledge in a number of areas: the history of relativization strategies, the effect of prescriptivism, the genre of scientific English and its textual organization, especially with respect to changes affecting the complexity of noun phrases, as well as American-British regional differences.

In section 2, we will briefly summarize the main findings of previous research regarding prescriptive grammar, regional differences in the use of relative clauses, overall diachronic developments, as well as findings on the use of relative clauses in scientific texts and their contribution to NP complexity. These studies provide the basis for our hypotheses. We focus on different types of relativizer and types of relative

[1] We would like to thank the anonymous reviewer for ELL for helpful comments on an earlier draft of the paper.

[2] Though the NP belongs to a PP in the case of pied piping (an instance of 'upward percolation' in the terminology of Huddleston, Pullum & Peterson (2002: 1040)).

clause but leave out the semantics of the antecedent. The data we use will be described in part 3 of our paper. In section 4, we briefly discuss analytical and theoretical problems related to different kinds of relative clause (adnominal vs. sentential, restrictive vs. non-restrictive) and the question of how a 'sentence' should be defined in historical texts. The results of our corpus analyses are discussed in section 5.

2 PREVIOUS RESEARCH

Since the focus in our paper is on historical data, written rather than spoken language use takes centre stage. Tagliamonte (2002: 163) suggests that "English is quite diglossic with respect to spoken and written norms at least with regard to the relativizer system". Where relevant, we will take variation between written and spoken English into account, but in the following review of earlier research we mostly focus on studies (especially in the area of historical developments) that have looked at written usage.

2.1 *Prescriptive tradition*

Sigley (1997) provides an excellent overview of the prescriptive tradition on relativizer choice. *That* with a personal antecedent, for instance, has a fairly complicated history:

> [it] was almost entirely displaced by *which* (at least in writing) by the late 17th century, but regained favour in time to be criticised by Addison (1711) [...]. In the meantime, the relative system had, through the spread of *who*, become newly organised as personal/impersonal, so that the arbiters of English were uncertain just where to put the reinstated *that.* (Sigley 1997: 72)[3]

[3] See also Fitzmaurice (2000: 199) on the codification of the *wh*-pronouns in eighteenth-century grammar.

While prescriptive opposition to *that* rather than a *wh*-pronoun in formal written language thus goes back to the eighteenth century, the prescriptive opposition to the use of *which* in restrictive relative clauses with an inanimate antecedent is a much more recent development. This is because the distinction between restrictive and non-restrictive relative clauses is recognised relatively late. In addition, restrictive relative clauses are the last environment in the spread of *wh*-pronouns (Sigley 1997: 72f.); so while Cobbett (1823: 28) allows for both *which* and *that* in restrictive relative clauses with inanimate antecedents, Bain (1863; cited in Morris 1895: 198) sees *that* as the only option (Sigley 1997: 73). After a preposition, *which* remains the only choice even in restrictive relative clauses.

Matters are further complicated by the fact that there is not a single prescriptive tradition that unifies 'approved' usage on both sides of the Atlantic: the British tradition targets non-restrictive *that*, whereas American arbiters of 'proper' English fight a war against the use of restrictive *which* (MWDEU: Gilman 1994: 895, see also Tottie 1997a: 86). The following comment in Taggart & Wines (2008: 141) illustrates the British prescriptive stand on non-restrictive *that:* "Non-restrictive relative clauses are introduced by the relative pronouns *who, whom, whose* and *which*, never by *that.*" On the other side of the Atlantic, a well-known example of the extreme opposition to restrictive *which* can be found in the influential style guide by Strunk & White (1999: 59)[4]:

[4] For more (and more varied) examples of recommendations in usage guides, college handbooks, in-house style guides at publishing houses and newspapers, etc., see Tottie (1997a: 85-7).

> The use of *which* for *that* is common in written and spoken language ("Let us now go even unto Bethlehem, and see this thing which is come to pass."). Occasionally *which* seems preferable to *that*, as in the sentence from the Bible. But it would be a convenience to all if these two pronouns were used with precision. Careful writers, watchful for small conveniences, go *which*-hunting, remove the defining *whiches*, and by so doing improve their work.

Some authors of usage guides seem to be aware of trans-Atlantic differences. Garner (2003: 782), for instance, puts the blame for the failure to use the relative pronouns 'correctly' squarely at the door of sloppy writers in the 'old' world:

> British writers have utterly bollixed the distinction between restrictive and nonrestrictive relative pronouns. Most commonly *which* encroaches on *that*'s territory, but sometimes too a nonrestrictive *which* remains unpunctuated.

In BrE usage, another distinction between the two relativizers takes the formality of the text into account. Fowler (1926: 635) criticizes the hypercorrect use of *which* in writing that results from this misconception:

> A supposed, & misleading, distinction is that *that* is the colloquial & *which* the literary relative. That is a false inference from an actual but misinterpreted fact; it is a fact that the proportion of *that*s to *which*s is far higher in speech than in writing; but the reason is not that the spoken *that*s are properly converted into written *which*s, but that the kind of clause properly begun with *which* is rare in speech with its short detached sentences, but very common in the more complex & continuous structure of writing, while the kind properly begun with *that* is equally necessary in both. This false inference, however, tends to verify itself by

> persuading the writers who follow rules of thumb actually to change the original *that* of their thoughts into a *which* for presentation in print.

2.2 *Regional differences*

The most comprehensive study on regional variation in relativizer choice is Sigley (1997).[5] On the basis of the Brown and LOB corpora and a parallel New Zealand corpus, he finds no significant differences between American and New Zealand academic or fictional writing; the only difference is that between American news language on the one hand and New Zealand as well as British journalese on the other hand (Sigley, 1997: 469): AmE prefers *that* over *which* as a subject relativizer in restrictive relative clauses. Sigley (1997: 114) also finds that in BrE and NZE, " the two relativizers *which* and *that* may be differentiated in terms of formality (…) rather than restrictiveness", thus confirming regional differences in the effect that prescriptive traditions may have had. Leech et al. (2009: 229-30) observe a marked difference in the choice of relativizers in the Brown family of corpora, namely a dramatic increase of relative clauses headed by *that* in American English, which is not paralleled in British English. They do not follow it up with a qualitative analysis of their data but speculate that the regional difference in this ongoing change is most likely due to the prescriptive rejection of restrictive *which* in the US:

---

[5] Note that Tottie (1997a) discusses differences in prescriptive stance on both sides of the Atlantic (a topic that is treated in more detail in Tottie 1997b); however, in her corpus analyses, she focuses on different relativizers and types of antecedent but does not distinguish between restrictive and non-restrictive relative clauses. For a study on relative clauses in some New Englishes, see Gut & Coronel (2012).

> Such a tradition has not been prevalent in usage guides in the UK, although since the early 1990s it has influenced countries throughout the world, including the UK, through its incorporation in internationally marketed word processors and grammar checkers. (Leech et al. 2009: 230)

The qualitative analysis of data in Hundt & Leech (forthcoming,2012) from the science section of the Brown family of corpora confirms the divergent development between AmE and BrE. Moreover, data from a more recent corpus of BrE texts sampled along the same lines as the Brown corpora suggest that BrE academic writing appears to be catching up with AmE in this area of usage (ibid.). In other words, grammar checkers do appear to have had a re-converging effect, with BrE following developments in AmE.

2.3 *Diachronic change in relativizer choice*

Previous literature on historical developments in relativizer choice is difficult to review because the studies tend to focus on different text types and regional varieties. More seriously still, they define the linguistic variable differently (e.g. only restrictive or both restrictive and non-restrictive; only adnominal or also sentential) and include different sets of relativizers (e.g. only overt relative pronouns or including zero).[6] The following overview can therefore only be a rough and necessarily incomplete sketch of a very complicated history.

Historically, zero and *that* are the older relativizers. The semantically more explicit *wh*-pronouns are introduced in the Early Middle English period (from learned

[6] See also Montgomery (1989: 114) and Ball (1996: 228) for a critique of existing research.

foreign models, see Mustanoja 1960: 110) and start spreading from the more formal to less formal written styles, especially in Early Modern English (see e.g. Dekeyser 1984: 65, Nevalainen 2002, Romaine 1980: 234). The *wh*-relativizers never become the dominant choice in informal and spoken English. Barber (1997: 213), on the basis of Elizabethan and Jacobean plays as well as Restoration Comedy, finds that "[t]he spread of *who* and *which*, and the recession of *that*, are especially characteristic of a formal style of writing. In informal and colloquial styles*, that* remains the commonest relative pronoun". Initially, *wh*-relativizers did not clearly differentiate between personal and non-personal antecedents (*which* could also be used with personal antecedents). According to Ball (1994, 1996), the semantic reorganization of the *wh*-relativizers along personal/impersonal lines occurred in the 17th century. In the late Modern period, *wh*-relative pronouns start impinging on the territory of *that* even in colloquial English. Grijzenhout (1992: 49) attributes this change to people's awareness of semantic differences:

> [...] by the year 1700 people became aware that *wh*-relatives have advantages which *that* does not have […]. This induced a change in the preference of *that* to one for *wh*-relatives in colloquial English which set in around the first decade of the eighteenth century.

On the basis of evidence from the Corpus of Nineteenth Century English (CONCE), Johansson (2006: 136f.) finds that *wh*-pronouns are used more widely than *that* in the nineteenth century. Furthermore "[i]n Science, the *wh*-forms are particularly frequent, occurring in 89 per cent of the cases" (2006: 137). The reason she gives is that "[t]he animacy and case contrasts signalled by the *wh*-forms […] contribute to the kind of clarity of expression and conciseness required of a scientific text" (2006: 137).

Ultimately, the popularity of *wh*-relatives in nineteenth-century scientific writing also means that they predominate in both restrictive and non-restrictive relative clauses (Johansson 2006: 145f.):

> The Science texts often contain logical reasoning, explanations and formulae: what is said in the preceding clause is expanded on in the next, and one step follows another. This is expressed in restrictive relative clauses, which occur in 80 per cent of the examples in this genre. Even if the relative clause is restrictive, *wh*-forms are used in more than 85 per cent of the cases. *Wh*-forms are typical of the formal scientific writing style as such, but they are also used because they convey the explicitness needed in a scientific text […].

In the twentieth century *that* increases again in written texts (see Leech et al. 2009: 227), a change that is spearheaded by American English (see previous section). In other words, relativizer choice in written texts shows a long-term development from *that* to *wh*-pronouns and a recent reversal of the trend towards a greater use of *that*.

This short account of the history of different relativizers simplifies the complexity of change, e.g. by not taking into account sentence length and distance between antecedent and relativizer (see e.g. Montgomery 1989, Rissanen 1984, Sigley 1997).

2.4 *Relative clauses in scientific English and NP complexity*

Apart from changes in relativization strategies, the development of the overall frequency of relative clauses has also been studied. However, genre-specific requirements with respect to formality and information packaging apply, and diachronic tendencies are therefore difficult to generalize to all genres. Different strategies in the

packaging of information (phrasal vs. clausal) bring us to the question of syntactic complexity. We are not concerned here with overall developments of syntactic complexity but with text-type-specific developments in the NP.[7] Douglas Biber in collaboration with various colleagues has looked at diachronic change in NP complexity across various text types. These studies provide a useful starting point for our own investigation.

Biber & Clark (2002: 63) measure complexity in the NP in terms of 'compression' and suggest the following cline for it:

---

[7] Biber & Clark (2002: 43) point out that there is little agreement on "the structural locus of complexity". A rather simplistic measure (sentence length and frequency of finite verbs) is used by Banks (2008: 67), even for the purpose of comparing different languages. Romaine (1980: 228f.) uses a measure of syntactic complexity that is based on Keenan & Comrie's accessibility hierarchy to contextualize the choice between *that* and *wh*-relatives. Recent work in Givón & Shibatani (2009) looks at the evolution of syntactic complexity from single words through phrases to clausal modification. In this article, however, we are mostly concerned with developments on the phrasal level rather than overall syntactic change; Pérez Guerra & Martínez Insua (2010a, b), who also study diachronic developments of phrasal complexity (albeit in the British letters and newspapers section of ARCHER rather than in scientific writing), not only take different types of pre- and postmodification into account but pay more attention to length of the modifier as well as internal complexity. Furthermore, they distinguish between different functions of the NP (subject, object). In terms of granularity of analysis, our study is more directly comparable with the work by Biber and colleagues.

COMPRESSED EXPRESSION	- premodifiers	< phrasal postmodifiers	< non-finite clauses	< relative clauses	- EXPANDED EXPRESSION

The compressed end of the cline is 'simpler' in terms of the number of elements and the overall length of the expression but from a cognitive perspective might be just as (if not more) complex. The expanded end of the cline, on the other hand, appears to be structurally more complex but in terms of processing – because it makes relations more explicit – could well be argued to be more accessible and thus 'simpler'. (For a discussion of 'complexity' from a typological, cognitive perspective, see Bisang (2009).)

Studies based on the British texts in ARCHER show that there has been diachronic shift (especially in twentieth-century informational writing) towards the more compressed end of expression, which goes hand in hand with less explicitness in meaning and thus greater decontextualization (Biber & Clark 2002: 68) as well as conceptual complexity. This fits in with previous research by Atkinson (1996, 1999) and Gotti (2003).[8] Surprisingly, however, the overall frequency of relative clauses seems to have remained relatively stable over time (Biber & Clark 2002: 57f., Biber & Gray 2011: 228f.); it is PPs that increase and thus it is a change in PPs that accounts for the difference in postmodification strategies in the twentieth century (Biber & Clark 2002: 59ff.).

[8] See also Gotti (2003: 83ff.) on the tendency of English specialized discourse to avoid subordination and to express conceptual complexity within the NP through nominalization and premodification rather than postmodification.

Biber & Conrad (2009: 164-5) make use of BrE medical writing in ARCHER. They describe the difference between eighteenth- and nineteenth-century research articles on the one hand and late twentieth-century scientific articles on the other hand as involving change from a clausal to a more nominal style:

> Science articles from earlier periods were mostly personal narratives of some kind or another.[9] As a result, these texts were composed of numerous clauses with a high density of verbs. [...] In contrast, modern research articles tend to use few verbs but numerous nouns and complex noun phrases.

This change is unlikely to be limited to medical writing. We expect to find similar tendencies in the science part of ARCHER. The research of Biber and his collaborators is also based on ARCHER, but only on the British part of the corpus. New data for American English has become available. This not only doubles the amount of available evidence but also allows us to add the dimension of regional variation to the picture. Biber et al. (2009) study modification in the NP on both sides of the Atlantic but only look at newspaper language. They find the same tendency towards more compressed NPs in this genre, too, but AmE is ahead of BrE in the development.

---

[9] A subtler characterization is given by Robert Sigley (p.c. 21 Feb. 2012), who asserts that "for most of the period represented in your data, the practice of science was conceived of in essentially Baconian terms: based primarily on the amassing of independent observations, with (e.g., causal) interpretation of those facts being deferred to a later stage (e.g. a later section of the text)". (Sigley was actually responding to another paper by two of the authors, in relation to the degree of relevance of a relative clause to a main clause and whether it might be marked off by punctuation.)

To sum up, there is no study so far that looks historically at regional variation in relativizer choice in both restrictive and non-restrictive relative clauses. Historical studies tend to look at genres or styles rather than compare regional varieties. We combine these two aspects in our study but limit our analysis to just one genre, scientific writing. In addition to choice of relativizer in different types of relative clause, we investigate the overall development of adnominal relatives vis à vis alternative modification strategies as an aspect of changing patterns of syntactic complexity within the NP in this specialised text type.

2.5 *Hypotheses*

On the basis of prescriptive traditions on both sides of the Atlantic and previous corpus-based research, we formulate the following hypotheses that we test against our corpus data:

1. Concerning relativizer choice

- Previous studies on the overall diachronic development in relativizer choice suggest that we should expect a shift from *which* to *that* in both varieties, even in scientific texts. This change will be visible in our American but not necessarily the British scientific texts.
- We expect *that* to be more frequently used in AmE: it is the relative pronoun actively advertised as the only grammatical option in restrictive relative clauses in this variety. British prescriptivists, on the other hand, target non-restrictive *that* as a variant to be avoided; an additional factor feeding a preference for *which* in BrE scientific writing is the opinion that it is the appropriate choice in formal written language.

2. Concerning relative clauses and change in NP structure

- Existing research into diachronic developments of NP complexity found a shift from clausal to phrasal modification as well as a shift from post-head to pre-head modification (see Biber & Clark 2002, Biber & Gray 2011, Biber, Grieve & Iberri-Shea 2009), which we also expect to find in our scientific data.
- A more compressed NP structure is likely to result in an overall decrease in sentence length. We therefore also investigate diachronic shifts along this parameter in our science texts.

3 CORPUS DATA AND METHODOLOGY

The material we use has been taken from ARCHER-3.2.[10] In addition to existing British English material we use American English scientific texts for all periods from 1700 onwards that were only recently added to the corpus. Table 1 gives an overview of the data.

[10] Collaboration and extension of the original ARCHER corpus has been going on for several years. For the development of the ARCHER corpus, see http://www.llc.manchester.ac.uk/research/projects/archer/ and Yáñez Bouza (2011).

	1700-49	1750-99	1800-49	1850-99	1900-49	1950-99
AmE	0	20,664	20,815	21,326	20,963	25,610
BrE	20,780	20,565	20,994	21,715	21,337	21,308

Table 1: Science texts in ARCHER-3.2 (number of words per sub-period)[11]

Furthermore, the science part of the ARCHER corpus was annotated with a parser (Pro3Gres) developed by Schneider (2008). Relative clauses were retrieved automatically from this syntactically annotated corpus. We discuss methodological issues (i.e. questions related to precision and recall) in a separate paper (Hundt, Denison & Schneider 2012). The parser was adapted after an initial run, and after parser adaptation, the recall for zero-, *that*- and *wh*-relatives was between 40% and 50% overall; precision was good at 82%-86% for *wh*- and *that*-relatives but quite poor for zero relatives. As part of the evaluation procedure, we analysed a subset of the corpus manually. We will also draw on these manually analysed sets of data for our analyses to test the validity of the results obtained on the basis of the automatically retrieved and post-edited sets of relative clauses.

4 Analytical and theoretical problems

4.1 *Adnominal vs. sentential relative clauses*

In the introduction we mention that we restricted our analysis to adnominal relative clauses. Real data are sometimes messy, so it comes as no surprise that some relative

[11] Our searches were based on a preliminary version of ARCHER-3.2 which includes two additional files for the second half of the twentieth century in the American subpart of the corpus, hence the slight imbalance in the size of subcorpora.

clauses defy easy classification as (a) adnominal vs. other, or (b) relative clause vs. complement clause. In the following example, for instance, the parser has wrongly identified *inspection* as the antecedent of a relative clause which in fact could either be postmodifying a pair of co-ordinated NPs or be attached to the preceding clause in a less specific way, in which case it would be sentential rather than adnominal:

(1) The practical outcome of this test is that arcs formed between these particular electrodes work most economically at from 1-8 to 2-2 kw. consumed in the arc itself; inspection of the curve showing that there is a marked falling-off of effectiveness below 1-8, and but very small increase above 2-2 kw., *added to which it was observed that higher powers caused the arc to burn unsteadily and to flare*, and in all probability caused the carbons to burn away with undue rapidity. (1925angu.s7b)

We discuss further problematic cases in Denison & Hundt (submitted). For the purposes of the present study, we manually excluded from our dataset all relative clauses that were not unambiguously adnominal.

4.2 *Restrictive vs. non-restrictive relative clauses*

The prescriptive ban on restrictive *which* is predicated on the notion of restrictive relative clause. A restrictive relative clause is one which serves to delimit the reference of the antecedent, to restrict it. Prescriptivists often maintain that the distinction is (relatively) unproblematic. Fowler (1926: 626), for instance, claims that "[t]here is no great difficulty […] about deciding whether a relative clause is defining [his term for 'restrictive'] or not; […]."

As a number of writers have pointed out, however, although a restrictive relative clause may be named from this logico-semantic function, the clause type has clear

syntactic and phonological correlates which are in many ways more central, such as that a restrictive relative clause forms a constituent with its antecedent, and that it belongs in the same intonation contour as the matrix clause. In scientific written data there is often a parenthetic interruption between antecedent and relative clause which makes the "phonological" test harder to carry out: one must imagine the written example edited down before being spoken aloud. The phonological property is in turn associated with the orthographic convention in writing of its not being marked off by commas. In historical data, punctuation is not a safe diagnostic, as many writers did not seem to punctuate reliably according to modern conventions (see Montgomery (1989: 137), who points out that punctuation of relative clauses only becomes standardized in the twentieth century, and Denison & Hundt (submitted), for developments in BrE scientific writing). In other words, a correlation between speech and punctuation cannot be relied on, especially in historical texts.

As has been pointed out (among others) by Lehmann (1984), Geisler & Johansson (2002), Sigley (1997) and Huddleston, Pullum & Peterson (2002), there are clauses which bear the distinctive formal signs of being "restrictive" relatives without being semantically restrictive; see Huddleston, Pullum & Peterson (2002: 1064-65). Conversely, non-restrictive relative clauses, usually regarded as supplying optional additional information, are sometimes effectively obligatory (Geisler & Johansson 2002: 96, citing Rydén 1984). Contrary to the prescriptivists' belief, the distinction is therefore a problematic one.

One solution, following Lehmann (1984), is to regard the distinction as gradient and to reclassify the dichotomy on the basis of the referential scope of the antecedent: generic vs. non-generic, and within the non-generic set, non-specific vs. specific vs.

unique. Another solution, adopted by Huddleston, Pullum & Peterson (2002: 1034-5; discussion 1058-66), is to retain a (recalibrated) dichotomy. In another paper we revisit the distinction, discuss alternative ways of classifying different types of relative clause and propose our own model (Denison & Hundt submitted). For the purposes of this paper we decided to retain the conventional dichotomy. For the majority of relative clauses automatically retrieved from the ARCHER science corpus, there was little or no doubt, but a number of examples were labelled as '?' on the first pass because the contextual evidence was not decisive. We then reviewed these queried examples in the light of the discussion in Huddleston, Pullum & Peterson (2002) to see whether the reinterpretation(s) they offer would resolve the uncertainty. In the end, whenever the balance of probability seemed to us clearly on one side or the other of the restrictive-nonrestrictive dichotomy, we simply counted that instance as unequivocal. We thus minimized the number of relative clauses initially analysed as 'unclear'. Examples (2) and (3) illustrate prototypical restrictive and non-restrictive relative clauses, respectively (note that neither of them is separated from the main clause by a comma):

(2) The comet was near two Stars *which are the 66th and 67th of Aquila and Antinous in the British Catalogue* […] (1724brad.s3b)

(3) Thus in the West I observ'd the Rays to be ting'd for some considerable time with an obscure and heavy Red; and in one of the brightest Streams at another time, there suddenly broke out a very vivid red *which was instantly and gradually succeeded by the other Prismatick Colours*, all vanishing in about a Second of Time. (1720cote.s3b)

4.3 *What is a sentence?*

In order to be able to discuss the development of relative clauses in relation to developments in NP complexity and possible repercussions for sentence length, we first need to define what we mean by 'sentence'. Crystal (2003: 414) claims that identifying sentences in written language is relatively straightforward, probably because punctuation is considered to be a helpful indicator of sentencehood. In modern written language, sentence boundaries are typically marked by full stops or exclamation/question marks. Once we start looking at historical data, however, the question as to what constitutes a sentence is not quite as straightforward because punctuation conventions seem to have undergone considerable change over time. In particular, use of the semicolon is much more frequent in historical data than in contemporary academic writing. Should semi-colons be added to the list of sentence boundary-markers? Table 2 shows the development over time:

sub-period	number of semicolons	number of sentences	semicolons per sentence
1700s	686	1553	0.447
1800s	616	2400	0.257
1900s	238	3532	0.067

Table 2: Semicolons per sentence in the science part of ARCHER (British and American subcorpora combined)[12]

[12] We would like to thank Paul Rayson (Lancaster University) for automatically annotating the corpus for sentence boundaries. The resulting files were not proofread, however. This produced some erroneous sentence analyses. The following is an example where two sentences were analysed as one (probably because S.W. was correctly tagged as an abbreviation): "[...] and in eight months out of the twelve, the least height of the barometer was accompanied with a S.W. This incited me to take the trouble of making out the preceding table, [...]." (1775hors.s4b)

The table shows that semicolons per sentence decrease substantially from the eighteenth to the twentieth century. We ended up deciding that only full stops, exclamation or question marks were to define sentence boundaries in our calculation of sentence length. The following is a typical example of a long sentence from our eighteenth-century data that provides us with an argument for excluding semicolons as sentence boundary markers.

(4) I took a Ball of Gold of an Inch in Diameter, that had a little Stem of the same Metal, with a place on it to fasten a String to; and having suspended it by a silken Thread too strong to lengthen by stretching, I made the Distance between the Center of the Ball, and the Point of Suspension equal to 12, 5 Inches, then causing the Ball to vibrate in a Trough full of Water, (which had an upright Piece of Wood in the middle of one side with Pins or Keys from which the Ball hung, that the Center of Suspension might always be in the same place) I observ'd by looking from a Pin on one side of the Trough to a mark made opposite to it on the other side, whereabouts the String of the Pendulum (just above the Surface of the Water; in which the Ball was quite immers'd) went after 14 Vibrations; and by another Pin and opposite mark, also observ'd where it went to, after 238 Vibrations. (1721desa.s3b)

There are three semicolons in this sentence. The first one could be replaced by a full stop. The second, however, precedes a relative clause that the parser had failed to identify because relative clauses after a semicolon were not included as a structural possibility in the parser grammar.[13] The third semicolon likewise precedes a sentence

[13] One might argue that relative clauses after a semicolon are more likely to be continuative relatives (for a discussion and definition of these, see Denison & Hundt submitted).

segment rather than a sequence that would result in a grammatical sentence were the semicolon to be replaced by a full stop.

5 FINDINGS

5.1 *Relativizers*

In section 5.1.1, we present results on the different types of relativizer that are used in our data, and look at regional as well as diachronic variation. We also compare the results from the automatically retrieved data sets with those from the manually analysed texts. In section 5.1.2 we look at the question of relativizer choice in different types of relative and the different prescriptive traditions in British and American English.

5.1.1 Overall developments in British and American scientific writing

In the American part of ARCHER, the dominant relativizer is *which*, particularly in the nineteenth century (see Figure 1). In the twentieth century, the proportion of *that* as a relativizer increases somewhat whereas zero relatives are used less frequently; *who* is also a low-frequency relativizer, a finding that most likely has to be attributed to the subject matter of scientific texts:

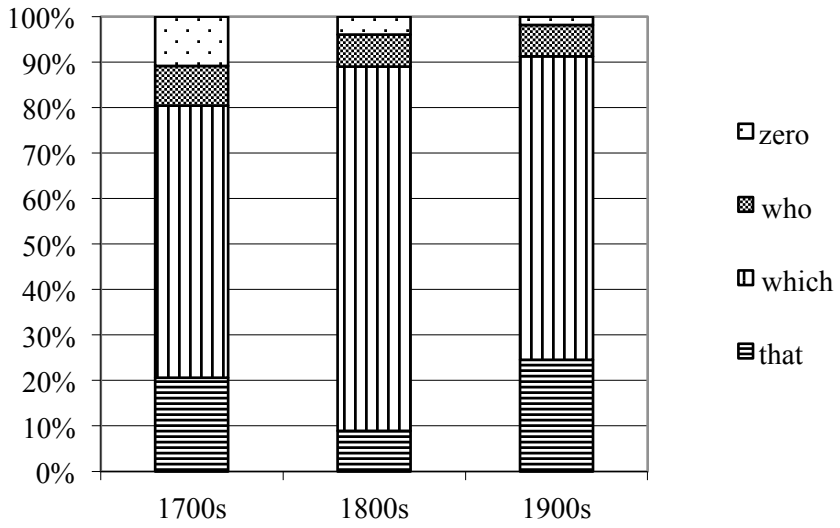


Figure 1: Relativizers (automatically retrieved and post-edited concordances of relative clauses) – AmE scientific texts (1700s, N = 184; 1800s, N = 200; 1900s, N = 285)

The main difference between our American and British data is that in the British data we see a steady decrease in *that*-relatives, whereas *which* rises to the position of dominant relativizer in the twentieth century (see Figure 2). Relative *that* is extremely rare in our BrE data. This probably has to be attributed to its being perceived as a spoken variant in Britain (see the comment by Fowler 1926: 635). This factor is likely to be stronger than the avoidance of *which* in restrictive relative clauses in Britain. The prescriptive stance on restrictive *which* in the US might account for the slightly lower proportion of this relativizer in our twentieth-century American data. We will take up this issue in the next section. Zero relatives, finally, show a more sudden decline in the British texts than in the American data.

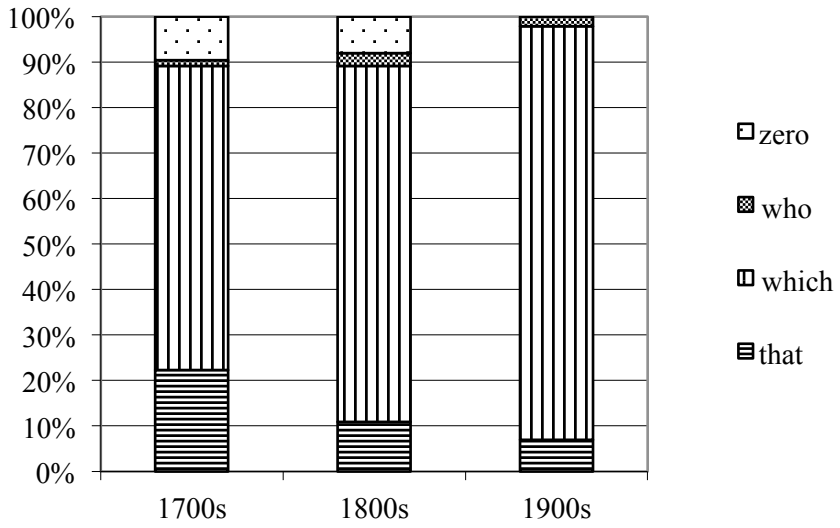


Figure 2: Relativizers (automatically retrieved and post-edited concordances of relative clauses) – BrE scientific texts (1700s, N = 286; 1800s, N = 274; 1900s, N = 144)

Before we look at the potential impact of prescriptive traditions, we would first like to see how the results obtained from the parsed data compare with those obtained from the manually analysed texts. We read both British and American texts for recall; the results on relativizer choice in the automatically retrieved data sets are collated in Figure 3a; Figure 3b gives the proportions of relativizers from the manually analysed texts. [14]

[14] We would like to thank Pius Meyer (University of Zürich) for reading some files for recall.

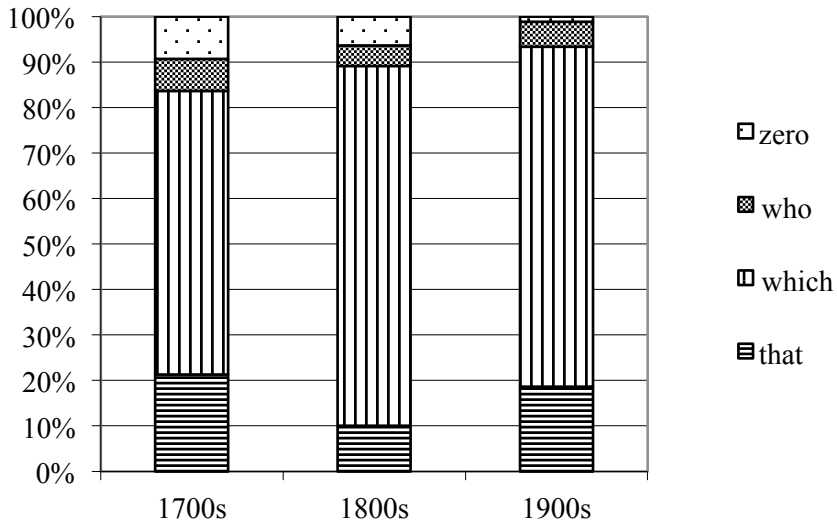


Figure 3a: Relativizers (automatically retrieved and post-edited concordances of relative clauses) – all of scientific texts (1700s, N = 470; 1800s, N = 474; 1900s, N = 429)

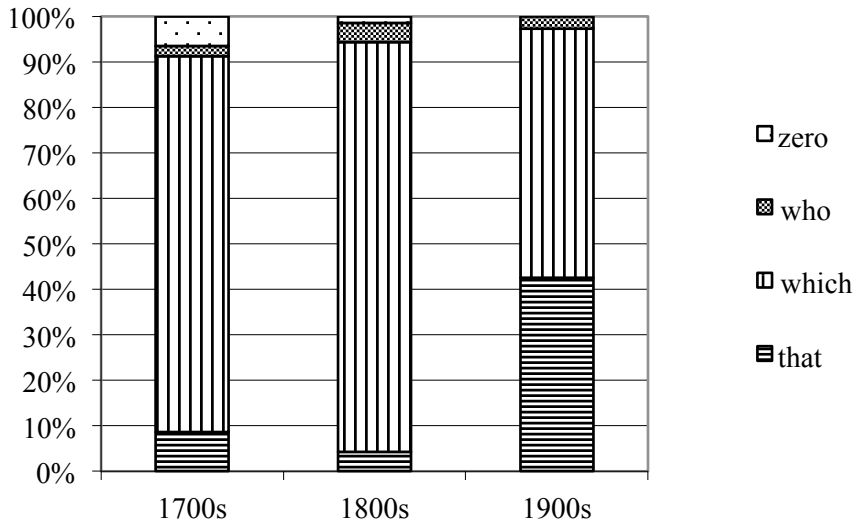


Figure 3b: Relativizers – results from manually analysed texts (1700s, N = 92; 1800s, N = 71; 1900s, N = 73)

The manually retrieved relative clauses yield a larger share of *that*-relatives only in the twentieth century. Overall, recall for *which* (in the automatically retrieved data) is lower than for relatives introduced by *that* in our scientific data (see Hundt, Denison & Schneider 2012). Thus, an important result that is confirmed by the comparative data from the manually analysed part of the corpus is that *which* is clearly the dominant relativizer. This finding is supported by evidence in Hundt (2011), who provides a manual analysis of late nineteenth- and early twentieth-century scientific texts in ARCHER: the automatic retrieval has much better recall for *that*- than for *which*-relatives. In other words, the automatically retrieved data give us a conservative picture with respect to the use of *which*-relatives in scientific English. The results reported in section 5.1.1 are therefore, on the whole, accurate with respect to the overall diachronic tendency, erring on the conservative side with respect to the dominance of *which* as relativizer in this text type. Were we to rely on manually retrieved data, the preference for *which* in scientific writing would be even more pronounced.

5.1.2 Relativizer choice and prescriptivism

In Figure 4 we present the results on types of relative (i.e. restrictive versus non-restrictive). They are calculated on the basis of all variable contexts, i.e. only those clauses with *wh*- or *that* as relativizer (zero can only introduce a restrictive relative clause). Furthermore, the analysis distinguishes between the American and the British part of the corpus because 'regional' variety is the relevant external variable that is of interest with respect to influence of prescriptivism. There is practically no change over time: restrictive relative clauses remain the most frequent type throughout (with somewhat more fluctuation in our British than American texts). This result fits in with

what Biber et al. (1999: 603) found in their investigation of relative clause types across genres: restrictive relative clauses are the most frequent kind in all types of writing (see also Peters 2004: 468).

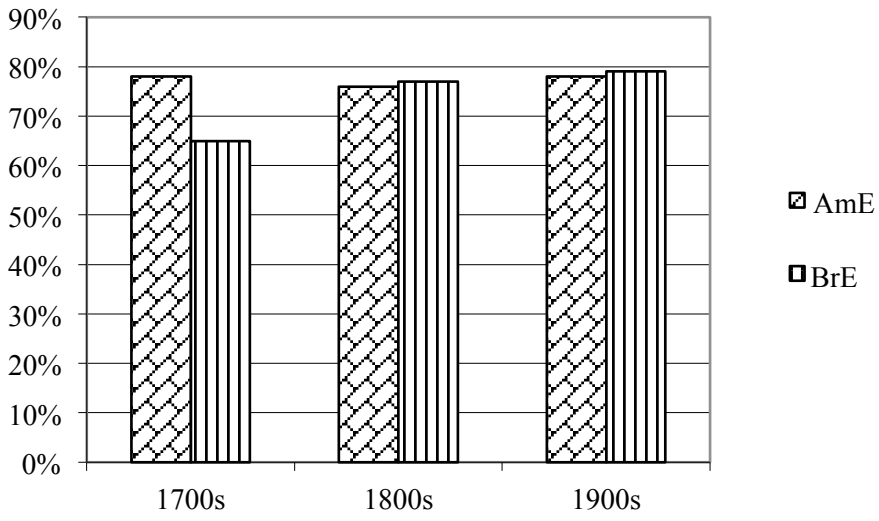


Figure 4: Proportion of restrictive relative clauses; automatically retrieved data (AmE 1700s, N = 164; 1800s, N = 192; 1900s, N = 210. BrE 1700s, N = 260; 1800s, N = 252; 1900s, N = 144)

As far as the distribution of relativizers in different types of relative clauses is concerned, our data support the hypothesis that, over time, American writers have become somewhat more prone to follow the prescriptive rule to use *that* in restrictive relative clauses rather than *which* (see Figure 5 below). But the results also show that despite the strong prescriptive tradition against restrictive *which* in the US, it is still the dominant relative pronoun in this type of relative clause in the twentieth century, at least in formal written usage (see also Sigley 1997: 414 on relativizer choice in academic writing in the twentieth century). In the BrE part of ARCHER, *which* clearly dominates in restrictive relative clauses.

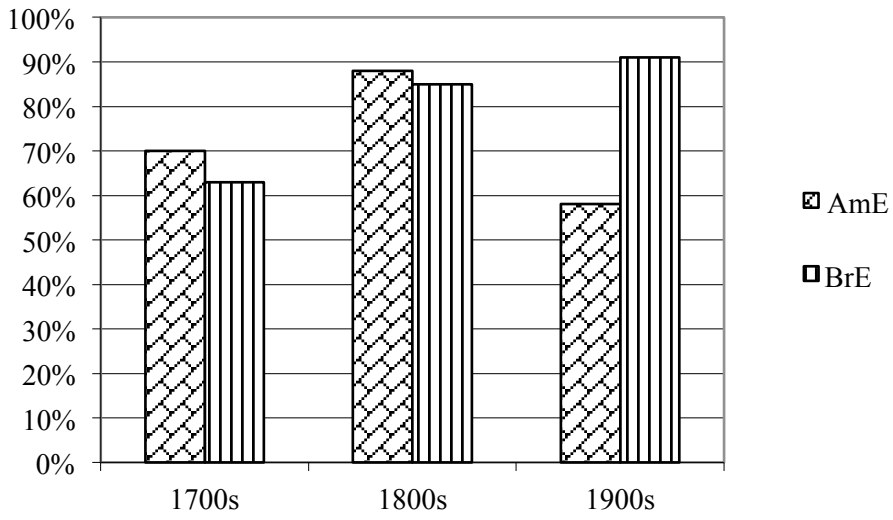


Figure 5: Proportion of *which* (vs. *that*) in restrictive relative clauses (AmE 1700s, N = 116; 1800s, N = 135; 1900s, N = 142. BrE 1700s, N = 157; 1800s, N = 188; 1900s, N = 113)

In non-restrictive relative clauses, on the other hand, we see that authors of scientific texts increasingly avoid *that* in our British data – though numbers were always low – and thus adhere to the prescriptive rule most commonly found in British style manuals (see Table 3). Non-restrictive *that* is also rare in AmE texts, but there is no diachronic trend to be observed.

	1700s		1800s		1900s	
	which	*that*	*which*	*that*	*which*	*that*
AmE	28	3	38	1	31	2
BrE	82	3	53	1	25	0

Table 3: *Which* vs. *that* in non-restrictive relative clauses

A possible example of a non-restrictive *that*-relative from our data is (5):

(5) I thought all my hopes of raising them [wild silkworms] were frustrated and concluded they would perish. I was agreeably surprized to see the little animals, *that I had given over as dead*, creeping out of their old skins, and appearing much larger and more beautiful than before. (1769bart.s4a)

5.2 *Relative clauses and NP complexity*

5.2.1 Sentence length

A look at the overall raw frequency of relatives clauses in our British and American English scientific texts shows that they decrease from 470 in the eighteenth to 429 in the twentieth century (see caption to figure 3a). At the same time, phrasal premodification increases, as we will show in section 5.2.2, resulting in a more compressed NP structure. This, in turn, is likely to be reflected in a decrease of overall sentence length. This assumption receives some support from Table 4:

	words	sentences	words per sentence
1700s	66,903	1,553	43.1
1800s	89,867	2,400	37.4
1900s	99,738	3,532	28.2

Table 4: Sentence length in scientific texts (BrE and AmE collated)[15]

Sentence length decreases somewhat from the 1700s to the 1800s, but a more marked decrease occurs towards the 1900s. This coincides with the marked decrease in relative clause frequency that we observe in our data. And of course a relative clause would increase the length of a sentence to which it was added more than a typical premodifier.

[15] Note that the number of words in this table are based on the parser counts rather than those given in table 1 above.

Furthermore, a decrease in sentence length corresponds to an increase in number of sentences pmw, so that the reduction in relative clause frequency must be even more striking on a per-sentence basis. Before we move on to other developments relating to the complexity of the NP in scientific discourse, let us briefly look at a couple of typical examples of long sentences from early academic writing. We already quoted an example of a long sentence from the eighteenth century in our discussion of the relation between punctuation and sentence boundaries. The following are good examples of the kind of long sentences found in nineteenth-century British academic writing:

(6) I now immediately arrived at that kind of general law *Ø I had been in search of*; for I found when things were thus arranged, that whatever might be the direction of the axis of rotation, if the motion of the ball were made towards the needle, the north end of the latter was attracted; and if from the needle, the north end was repelled by the iron, in points immediately in the axis (when of course the motion of the shell was parallel to the needle) being neutral, or those *at which the change of direction took place*; in other words, if the motion of the shell continue the same, and the compass be successively placed all round the ball, in that semi-circle (from one axis to the other) *in which the motion is towards the needle*, the north end approaches the ball, and in the other semicircle it recedes, or the south end approaches; the points of non action being in the two extremities of the axis , and those of maximum effect in two opposite points at right angles to the axis; *in which two latter the needle*, when properly neutralized, *points directly to the centre of the ball*. (1825barl.s5b)

(7) Thus a sheet of copper 4 feet long, 14 inches wide, and weighing 9 lb. 6 oz., protected by 1/100 of its surface of cast iron gained in ten weeks and five days, 12 drachms, and was coated over with carbonate of lime and magnesia: a sheet of copper of the same size protected by 1/150, gained only 1 drachm in the same time, and a part of it was green from the adhering salts of copper; whilst an unprotected sheet of the same class,

> both as to size and weight, and exposed for the same time, and as nearly as possible under the same circumstances, had lost 14 drachms; but experiments of this kind, though they agree when carried on under precisely similar circumstances, must of necessity be very irregular in their results, when made in different seas and situations, being influenced by the degree of saltness, and the nature of the impregnations of the water, the strength of tide and of the waves, the temperature, &c. (1825davy.s5b)

Example (8) shows that there is some residual evidence of longish sentences to be found even in twentieth-century academic writing:

(8) The cost of producing a given effect is the product of the energy and the time *for which this energy is maintained*, and it was hoped that by multiplying each applied power in kilowatts by the number of minutes *which it took to kill the infusoria*, the kilowatt-minutes required for a lethal dose thus obtained, plotted against the energy in kilowatts for each dose, would give a regular curve showing a minimum value of kilowatt-minutes, for some critical value of power, or one *from which such a minimum might be calculated.* (1925angu.s7b)

Interestingly, this sentence contains three relative clauses. In addition, it contains postmodifying participle clauses introduced by past participles (*required, plotted*) or a present participle (*showing*). We will return to these types of clause below.

5.2.2 Phrasal premodification

As pointed out above, complexity of the NP can be achieved by non-clausal means, resulting in a more compressed (and thus cognitively more complex) structure. In example (9), a complex ADJ phrase (that could easily be turned into a non-restrictive relative clause) postmodifies the head; examples (10) and (11) contain postmodifying PPs which could likewise be expanded into relative clauses:

(9) The youngest soils (No. 4 in Table 2), *more or less correlative with the pottery cultures*, have a weakly developed leached zone [...] (1955hunt.s8a)

(10) Some podsolic soils *with well-developed leached zones* are prepottery in age [....] (1955hunt.s8a)

(11) The limited time *at a field worker's disposal* and his desire to cover as broad a range of phenomena as possible often lead him to associate with persons in the community who are congenial in the sense of accepting him and giving him information. (1954honi.s7a)

So far, the examples we have discussed are all of post-head modification. NP complexity can also be achieved by multiple pre-head modification, either with adjectives (12) or nouns (13); examples in (14) show how both types of premodification easily combine in complex NPs.

(12) a. The intense short rays (1925angu.s7b)

b. the chief spherical harmonic terms (1925cha1.s7b)

c. Magnetic field-induced orientation (1975duru.s8b)

(13) a. a Constant Water Vapour Addition (1925fenn.s7b)

b. Barapasaurus gen. nov. Derivation (1975jain.s8b)

c. an earthquake ground fracture (1975tcha.s8b)

d. Prof. E. W. MAcBRIDE (1925gord.s7b)

(14) a. the other basic hydrolysis products

b. no corresponding large pressure differences (1975crap.s8b)

c. the average effective stress level (1975bish.s8b)

In addition to the development of relative clauses, we therefore also investigated the development of other types of post- and premodification pattern.

Figures 6 and 7 show that pre-head modification with nouns or adjectives increases towards the twentieth century, a development that, overall, is more

pronounced in AmE than in BrE (Leech et al. 2009: 216f.).[16] The results in figures 6 and 7 are even more striking if we take into account that the NPs were retrieved automatically from our data and that the evaluation of precision shows that the datasets from the 1700s and 1800s contain more false positives than those from the 1900s (see table 5 below).

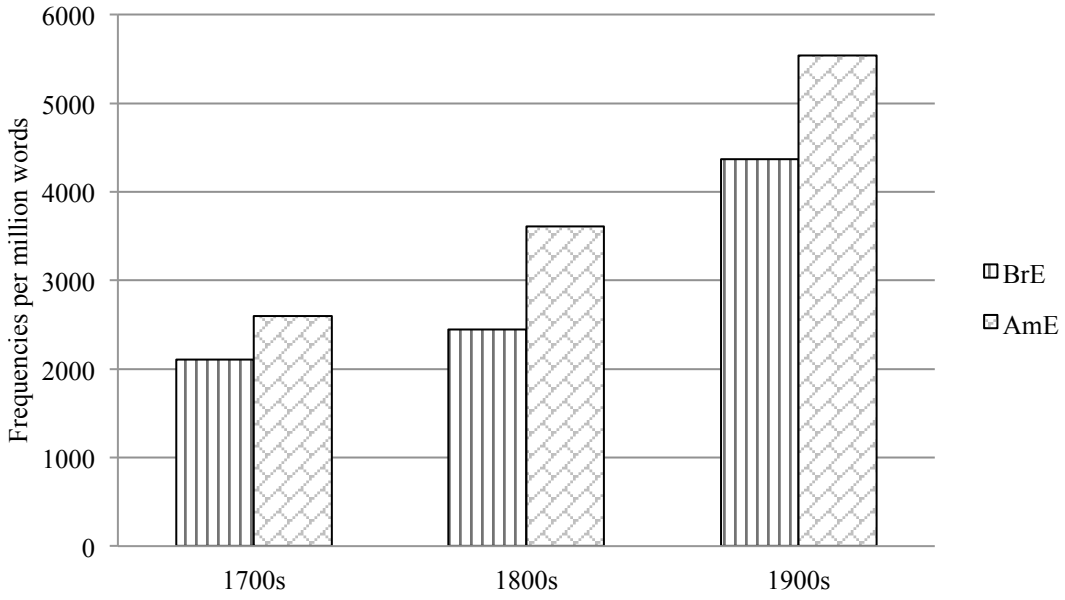


Figure 6: adj-adj sequences in the science sub-corpus of ARCHER

[16] *S*-genitives were excluded from the counts. Note that we give the results as constructions pmw. An alternative measure would be to calculate the relative frequency per NP, in case differences in the development of different parts of speech over time added 'noise' to the statistics. The parsed data allow us to calculate per noun chunk, but it turns out that the same overall trend emerges from the differently calculated measure (see tables 1a and 1b in the appendix).

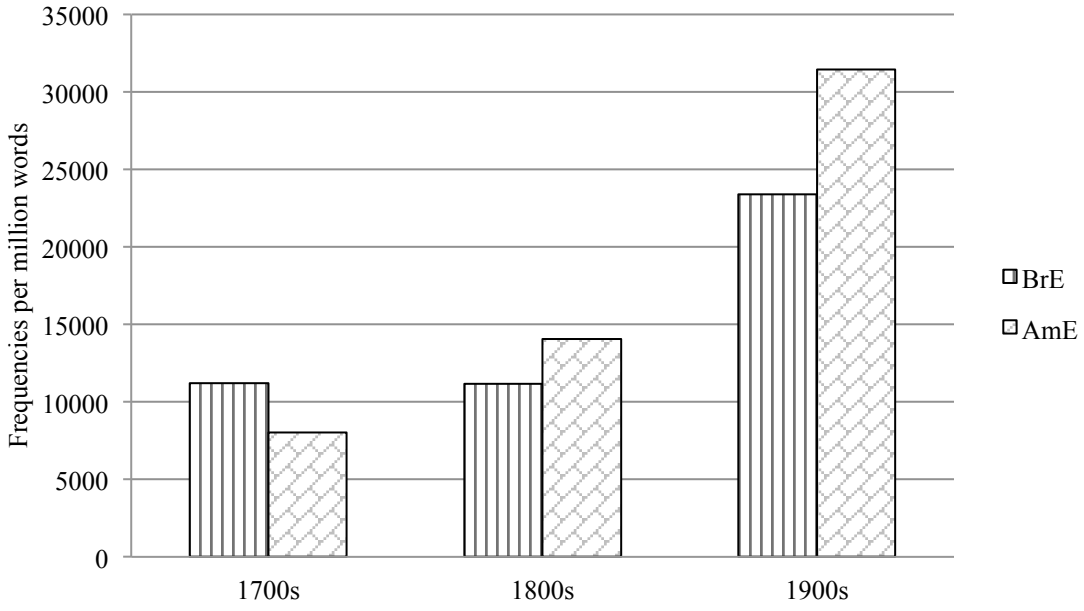


Figure 7a: NN sequences in the science sub-corpus of ARCHER

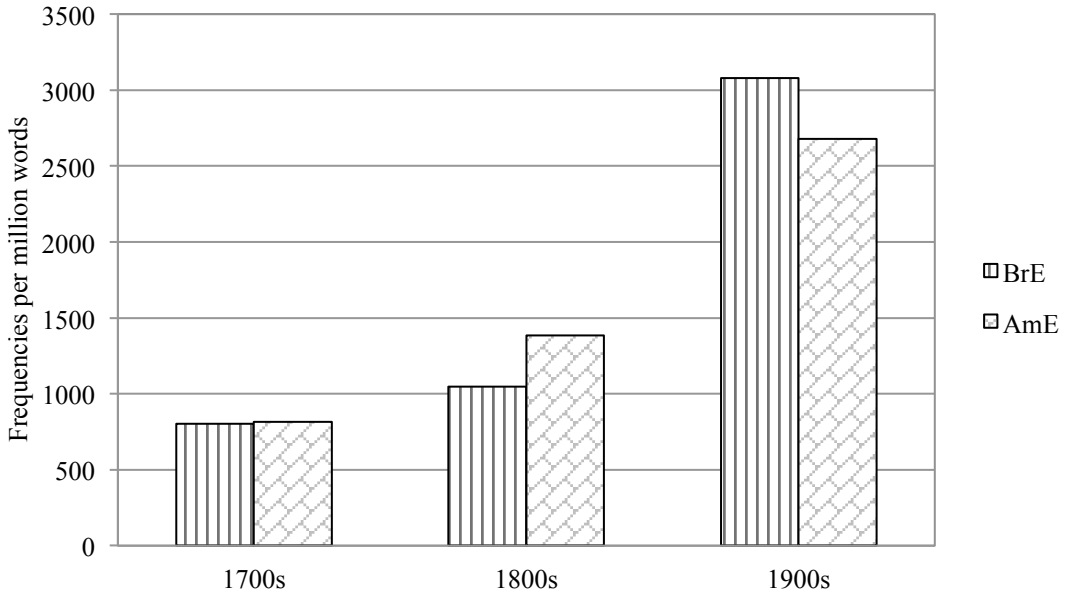


Figure 7b: NNN sequences in the science sub-corpus of ARCHER

As illustrated in (13) d. above, the data on which figures 7a and 7b are based include instances with proper names as heads. Biber & Gray (2011: 237) point out that examples prior to 1800 were proper names with multiple titles; sequences of nouns that are not proper names start occurring only after 1800 in their data. We therefore also

searched for combinations of nouns that modify a common noun rather than a proper name. The results in figures 8a and 8b illustrate the same overall trend.

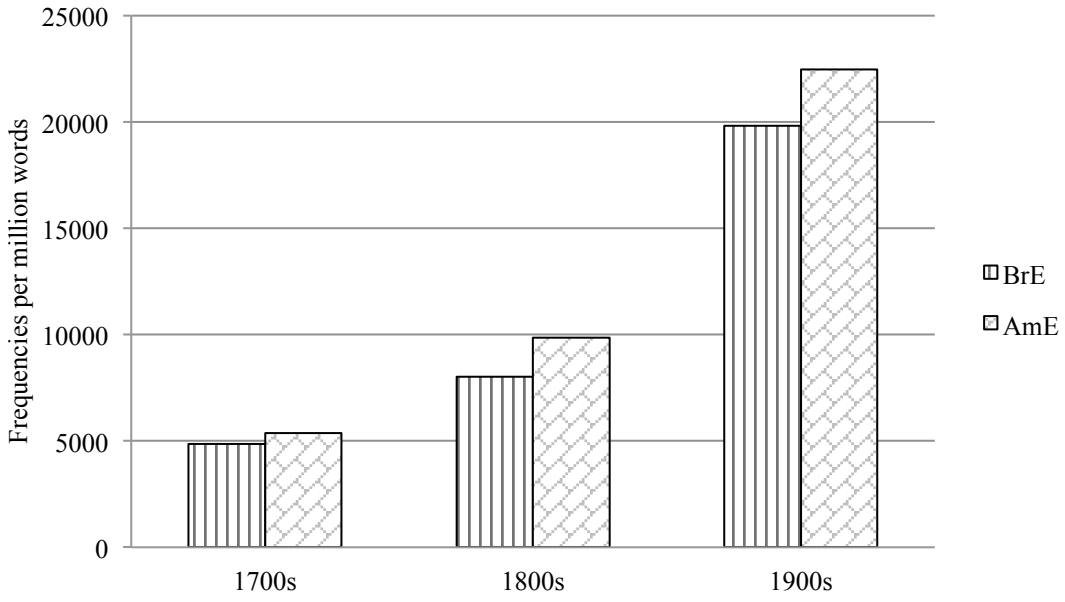


Figure 8a: NN sequences in the science sub-corpus of ARCHER (excluding proper name as head)

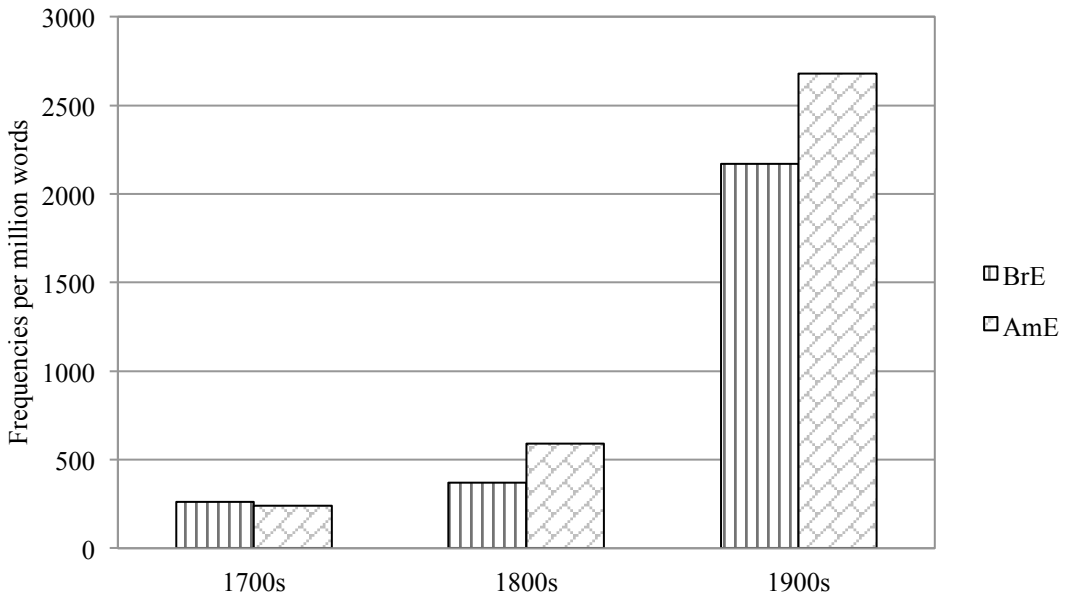


Figure 8a: NNN sequences in the science sub-corpus of ARCHER (excluding proper name as head)

Moreover, on closer inspection, early examples from the 1700s turn out to be Latin nouns such as in *the Fluxus menstruus immodicus* (1720perc.s3b), or parser errors. The

first undisputed NNN sequences come from a 1791 article: *the Sugar Maple tree* and *the sugar maple country* (1791rush.s4a), but these arguably contain compound nouns and might therefore not classify as prototypical NNN sequences. The following illustrate the first genuine sequences of common nouns that are variants of noun phrases which could have been postmodified by a clause or PP:

(15) a. the internal-combustion engine standpoint (1925fenn.s7b)

vs. the standpoint of the internal-combustion engine

b. the induced pore water tension (1975bish.s8b)

vs. tension of pore water induced by…

c. interspecific pollen tube growth inhibition (1975hoge.s8b)

vs. interspecifically inhibiting the growth of the pollen tube

Furthermore, these 'true' NN and NNN sequences increase in the twentieth century. In the scientific texts from ARCHER (BrE and AmE collated), there are 590 NNN-sequences per million words in the 1800s. In the first half of the twentieth century, they have increased to 1662 pmw (N =78); figures almost double again to 3030 pmw in the second half of the century (N=160). Our study thus confirms Biber & Gray's (2011: 238) findings on these constructions in BrE medical writing: "The dramatic change in use for these structures occurred in the second half of the twentieth century, when NNN sequences become relatively common, and even NNNN sequences are not unusual." Moreover, in their qualitative analyses they found that semantic relationships between the nouns expand over time (Biber & Gray 2011: 238-40). In other words, there is not just a change in frequency but also one in function: "… the grammatical features themselves have undergone major extension in their lexical associations, grammatical variants and functions, and meanings" (Biber & Gray 2011: 248).

5.2.3 Clausal postmodification

Biber et al. (2009) only look at phrasal modification and relative clauses. They mention other types of clausal modification (e.g. *to*-infinitives, participle clauses) as variants in their study of NP complexity, but they do not provide any quantitative evidence on their development. The reason for this is most likely that they use a tagged corpus, and participle clauses are virtually impossible to extract from a tagged-only corpus. Our parsed data allow us to extract this information. As figure 9 shows, clausal postmodification with participle clauses also increases over time. Again, the diachronic trend is clearer in AmE texts than in BrE scientific writing.

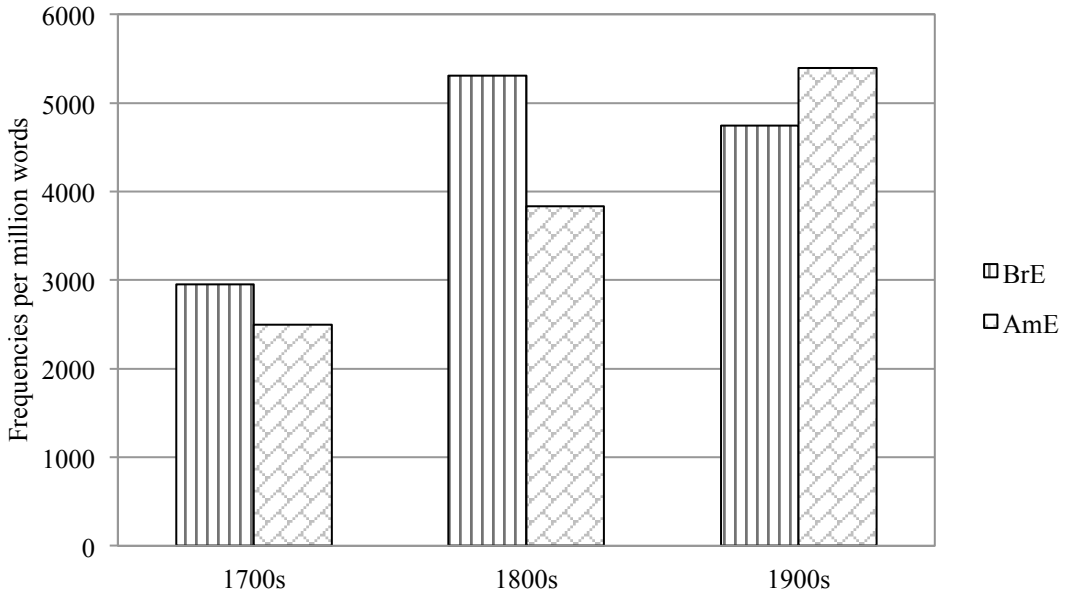


Figure 9: Postmodifying participle clause (*-ing/-ed*) in the science sub-corpus of ARCHER

If we look at the two types of non-finite postmodifying clause separately, we see that BrE is initially more advanced in using *-ing* clauses, but AmE takes the lead in the twentieth century (see figure 10a); the peak for participle clauses in the 1800s BrE part

of ARCHER clearly has to be attributed to clauses introduced by a past participle (see figure 10b).

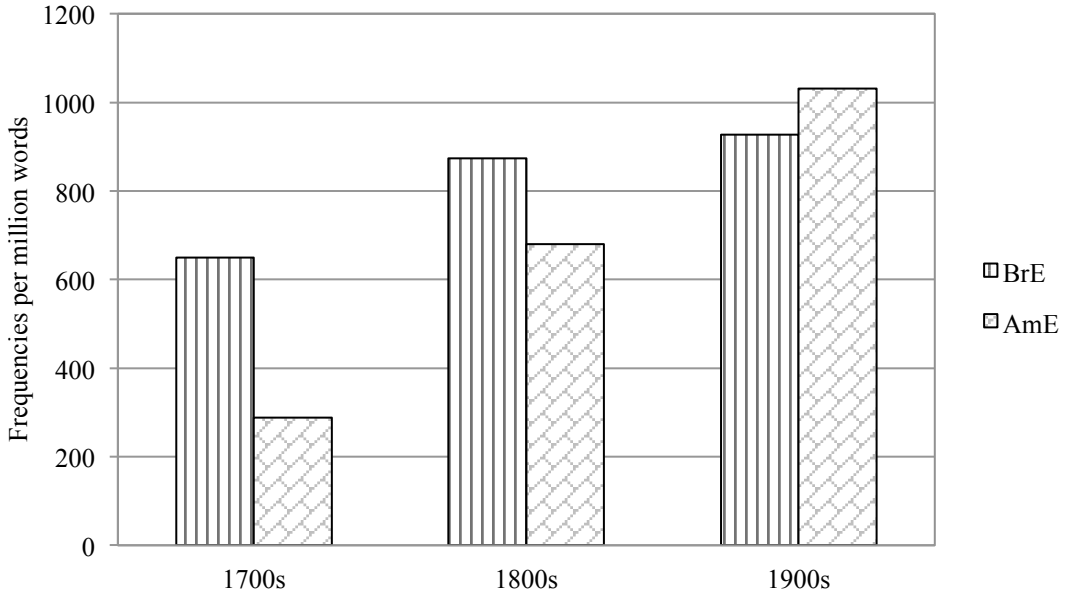


Figure 10a: Postmodifying *-ing* clauses in the science sub-corpus of ARCHER

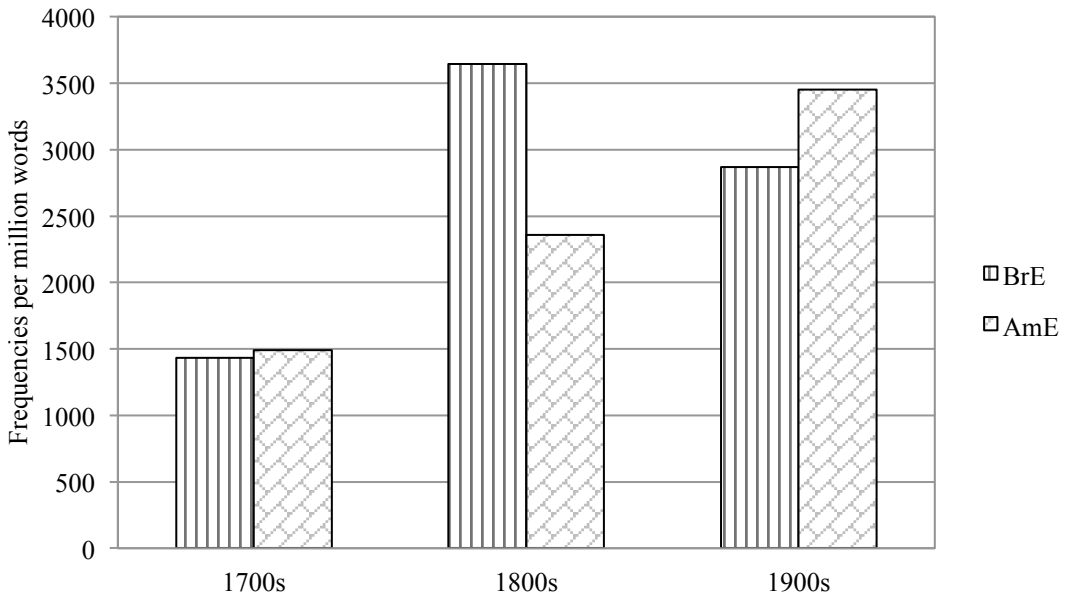


Figure 10b: Postmodifying *-ed* clauses in the science sub-corpus of ARCHER

Postmodifying participle clauses would potentially be reduced relative clauses, but this is a fuzzy category (see Hundt, Denison & Schneider 2012 for more detailed

discussion), and we therefore refrain from labelling them as such. Regardless of their theoretical status, participle clauses are of particular interest in our study for the following reason. Our evidence on participle clauses adds a new twist to the story of NP complexity – not only does premodification (N, NNN, adj-adj) increase over time, but there also seems to be a trade-off between overt relativization and participle clauses (candidates for reduced relative clauses), as Figure 11 shows. In the 1900s scientific part of ARCHER, participle clauses are more frequent than relative clauses. This development is more obvious in the American sub-corpus than in the British one (see Figures 12a and 12b). Participle clauses provide a slightly denser form of information packaging than overt relative clauses, but the resulting NPs are not quite as 'compressed' as those with phrasal modification.

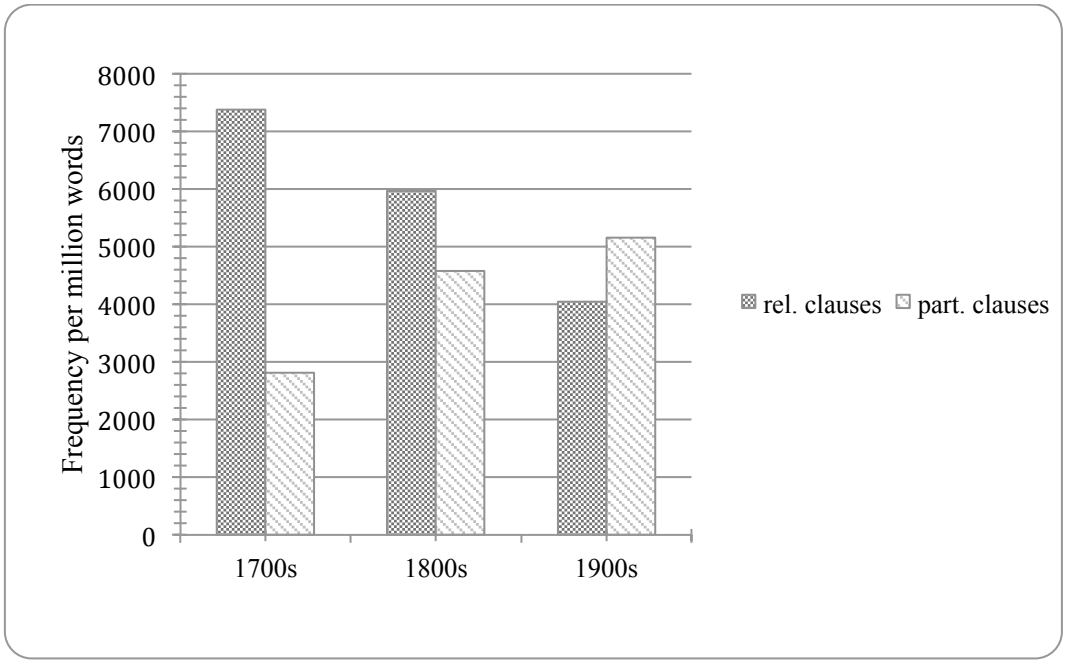


Figure 11: Development of clausal postmodification (relative clauses vs. participial clauses; BrE and AmE scientific texts combined)

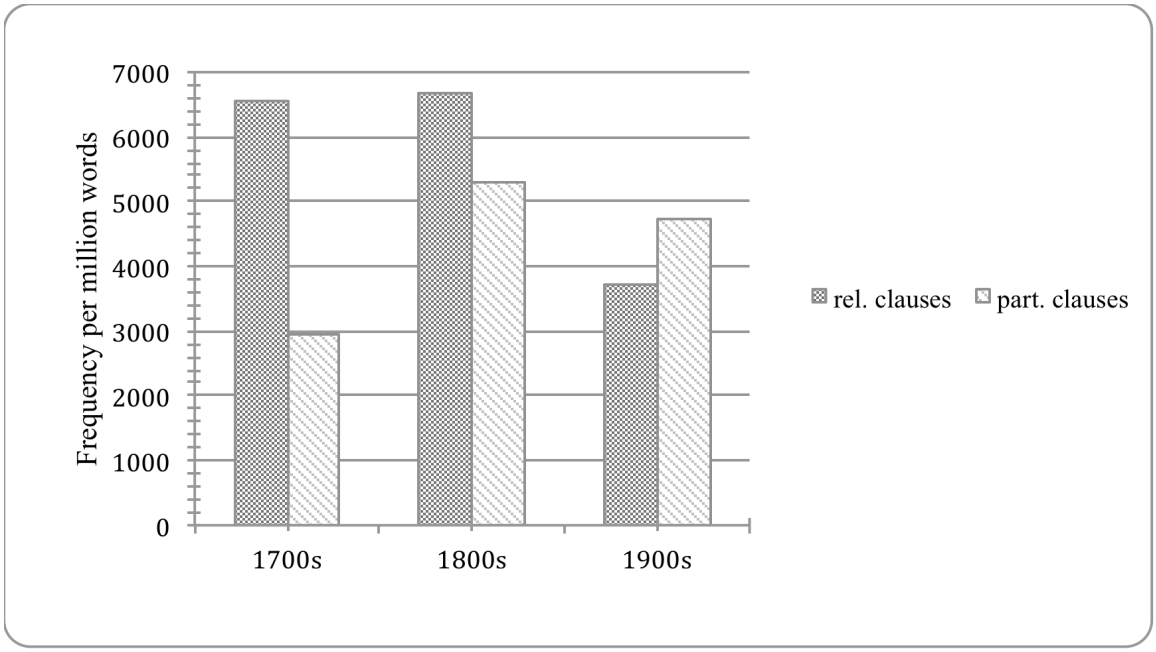


Figure 12a: Development of clausal postmodification (relative clauses vs. participial clauses (BrE scientific texts)

Figure 12b: Development of clausal postmodification (relative clauses vs. participial clauses (AmE scientific texts)

5.2.4 Evaluation of automatically retrieved data

The data for pre- and postmodification were extracted from the parsed corpus but not manually post-edited. However, we evaluated the precision of the parser (as well as tagger and chunker errors leading to parser errors) on the noun phrase complexity features described in figures 6 to 12. For each structure and each century, we manually verified the output of 100 random sentences (or all sentences, if counts were below 100). The percentages are given in Table 5.

	1700s	1800s	1900s
adj-adj sequences (figure 6)	89%	98%	95%
NN-sequences (figure 7a)	64%	76%	94%
NNN-sequences (figure 7b)	88%	79%	89%
NN-sequences excluding proper names (figure 8a)	62%	82%	91%
NNN-sequences (excluding proper names figure 8b)	62%	58%	78%
postmodifying –ing clauses (figure 10a)	89%	84%	79%
postmodifying –ed clauses (figure 10b)	80%	78%	84%
overt relative clauses (figure 11)[17]	86%	83%	86%

Table 5. Precision evaluation on noun complexity structures

As a trend, parser performance is lower on historical data. Precision for nouns is affected more seriously, as 'noun' is a default tag for unknown words. This partly explains the low performance of the parser on the historical texts in figures 7 and 8. In general, the precision of the parser-based data is high enough to confirm the developments described in sections 5.2.1-5.2.3 above.

[17] For a more detailed discussion of precision and recall of automatically retrieved relative clauses, see Hundt, Denison & Schneider (2012).

6 SUMMARY AND CONCLUSION

With respect to relativizer choice, our study confirms that *that* is used more frequently in American scientific writing than in the corresponding British part of ARCHER. Contrary to the developments predicted in previous literature, there is no shift from *which* to *that* in our data. In the British part of the corpus, *that* shows a steady decline from the 1700s to the 1900s; in AmE it decreases and then increases again, but just slightly beyond its original frequency in the 1700s. The dominant relativizer in both varieties is *which*. To some extent, this might have to do with the more transparent semantics of the *wh*-relatives or their perceived formality. Surprisingly, however, *which* is still the dominant relative pronoun even in restrictive relative clauses on both sides of the Atlantic (this holds both for our automatically retrieved datasets as well as the manually retrieved relative clauses). In other words, the American war on restrictive *which* is not reflected in our data. The success of prescriptive influence on relativizer choice in the US (see Hundt & Leech forthcoming,2012, Leech et al. 2009)) therefore turns out to be a fairly recent development. The British prescriptive stance on the avoidance of *that* as an informal variant, on the other hand, finds support in our corpus results. Overall, restrictive relative clauses are the most frequent type across time and variety. Our results confirm previous studies on this (e.g. Biber et al. 1999, Johansson 2006).

With respect to NP complexity, we found that the frequency of relative clauses decreases in both BrE and AmE scientific writing (see Biber & Clark 2002, Biber & Gray 2011). At the same time, we see an increase in some kinds of premodification (i.e. AAN-, NN- and NNN-sequences and combinations thereof). This supports previous findings on a growing densification of the noun phrase in informational writing. This

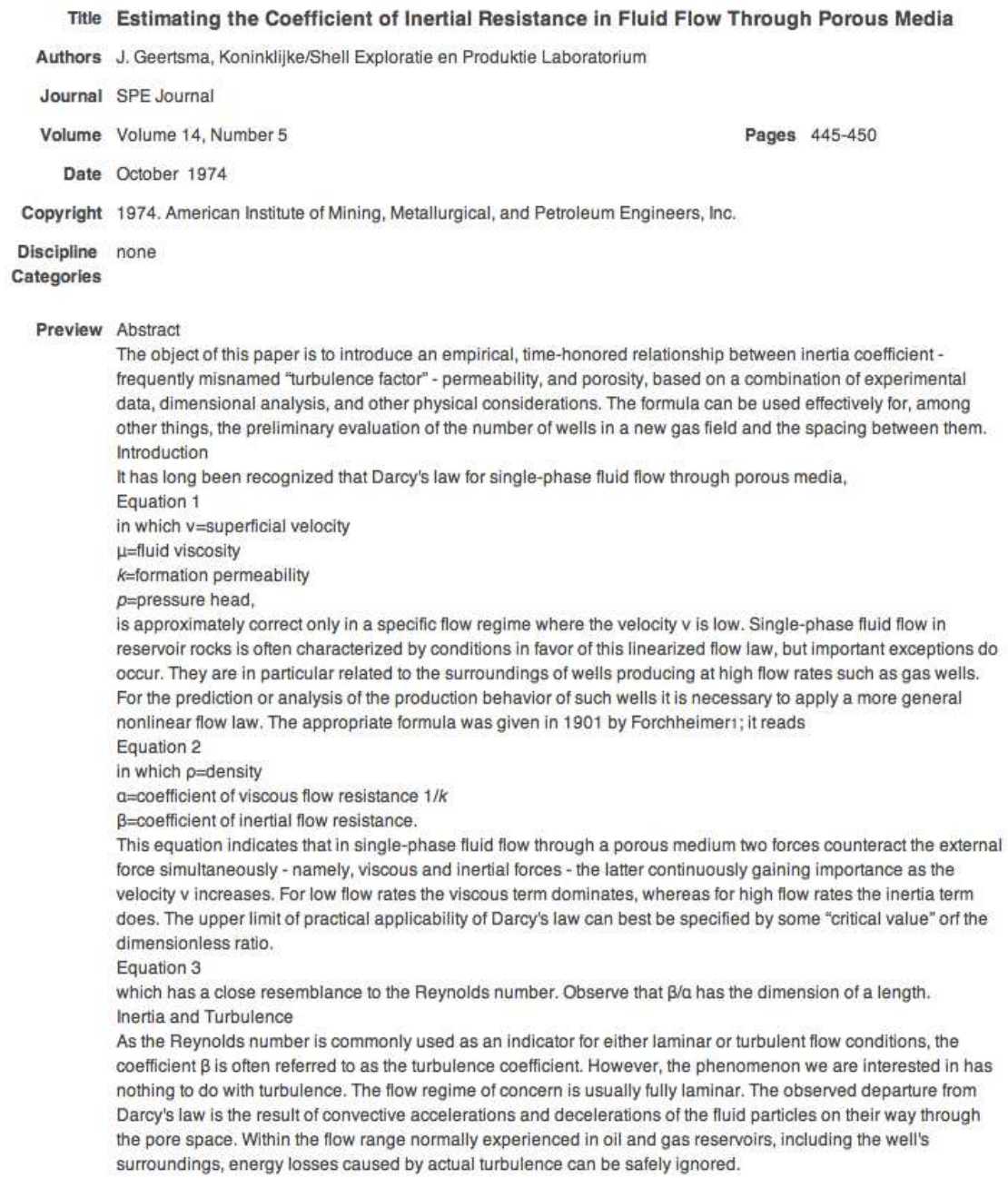
trend has repercussions in the development of overall sentence length, which decreases over time. The diachronic shift to more compressed noun phrases is also evidenced on a slightly less spectacular level: there seems to be a trade-off between relative clauses (decrease) and postmodifying participle clauses (increase). In other words, a slightly less expanded form of clausal postmodification increases at the expense of a more expanded one. If different types of clausal modification are taken into consideration, the shift from clausal to phrasal modification (in scientific English) appears to be a little less marked than previously claimed. But the overall trend is definitely from more expanded to less expanded.

The question is why we should see such changes and how we are to interpret them. One answer can be found in the development of the text type. In terms of text type functions, (Biber & Conrad 2009: 166) point out that "… science research articles have shifted in their specific purposes, and they have become much more narrowly defined in terms of textual conventions, but throughout they have maintained the basic communicative goal of conveying the results of scientific inquiry". However, with the 'informational explosion' in the twentieth century, the pressure to communicate information efficiently has increased (see Biber & Clark 2002: 63f., Biber & Gray 2011: 234f.). Figures 13 and 14 show the opening passages of an eighteenth- and a twenty-first century article on a related topic, the investigation of resistance in fluids, that serve to illustrate the developments from a more involved, personal style of scientific writing to a more impersonal/informational one.

IV. *Experiments relating to the Reſiſtance of Fluids, made before the* Royal Society *on* Thurſday, March *the* 30*th*, 1721. *By the Reverend* J. T. Deſaguliers, LL. D. *F. R. S.*

I Took a Ball of Gold of an Inch in Diameter, that had a little Stem of the ſame Metal, with a place on it to faſten a String to; and having ſuſpended it by a ſilken Thread too ſtrong to lengthen by ſtretching, I made the Diſtance between the Center of the Ball, and the Point of Suſpenſion equal to 12,5 Inches, then cauſing the Ball to vibrate in a Trough full of Water, (which had an upright Piece of Wood in the middle of one ſide with Pins or Keys from which the Ball hung, that the Center of Suſpenſion might always be in the ſame place) I obſerv'd by looking from a Pin on one ſide of the Trough to a mark made oppoſite to it on the other ſide, whereabouts the String of the *Pendulum* (juſt above the Surface of the Water; in which the Ball was quite immers'd) went after 14 Vibrations; and by another Pin and oppoſite mark, alſo obſerv'd where it went to, after 28 Vibrations. Taking out the Water, I fill'd the Trough with Mercury, the length of the *Pendulum*, Point of Suſpenſion and all other things remaining as before: then letting go the Ball in the Mercury from the ſame place whence it was let down when the Trough was full of Water; (which was mark'd by a String ſtretched a croſs to prevent miſtakes) after one

Figure 13: Opening passage of an eighteenth-century research article (*Philosophical Transactions of the Royal Society*, Vol. XXXI)



Title **Estimating the Coefficient of Inertial Resistance in Fluid Flow Through Porous Media**

Authors J. Geertsma, Koninklijke/Shell Exploratie en Produktie Laboratorium

Journal SPE Journal

Volume Volume 14, Number 5 **Pages** 445-450

Date October 1974

Copyright 1974. American Institute of Mining, Metallurgical, and Petroleum Engineers, Inc.

Discipline Categories none

Preview Abstract

The object of this paper is to introduce an empirical, time-honored relationship between inertia coefficient - frequently misnamed "turbulence factor" - permeability, and porosity, based on a combination of experimental data, dimensional analysis, and other physical considerations. The formula can be used effectively for, among other things, the preliminary evaluation of the number of wells in a new gas field and the spacing between them.

Introduction

It has long been recognized that Darcy's law for single-phase fluid flow through porous media,

Equation 1

in which v=superficial velocity

μ=fluid viscosity

k=formation permeability

p=pressure head,

is approximately correct only in a specific flow regime where the velocity v is low. Single-phase fluid flow in reservoir rocks is often characterized by conditions in favor of this linearized flow law, but important exceptions do occur. They are in particular related to the surroundings of wells producing at high flow rates such as gas wells. For the prediction or analysis of the production behavior of such wells it is necessary to apply a more general nonlinear flow law. The appropriate formula was given in 1901 by Forchheimer1; it reads

Equation 2

in which ρ=density

α=coefficient of viscous flow resistance 1/*k*

β=coefficient of inertial flow resistance.

This equation indicates that in single-phase fluid flow through a porous medium two forces counteract the external force simultaneously - namely, viscous and inertial forces - the latter continuously gaining importance as the velocity v increases. For low flow rates the viscous term dominates, whereas for high flow rates the inertia term does. The upper limit of practical applicability of Darcy's law can best be specified by some "critical value" orf the dimensionless ratio.

Equation 3

which has a close resemblance to the Reynolds number. Observe that β/α has the dimension of a length.

Inertia and Turbulence

As the Reynolds number is commonly used as an indicator for either laminar or turbulent flow conditions, the coefficient β is often referred to as the turbulence coefficient. However, the phenomenon we are interested in has nothing to do with turbulence. The flow regime of concern is usually fully laminar. The observed departure from Darcy's law is the result of convective accelerations and decelerations of the fluid particles on their way through the pore space. Within the flow range normally experienced in oil and gas reservoirs, including the well's surroundings, energy losses caused by actual turbulence can be safely ignored.

Figure 14: Opening passage of a twenty-first-century research article (http://www.onepetro.org)

The example in figure 13 already shows a development from the earlier epistolary format of research 'articles' in that it opens with a title as well as a reference to the place and time where the paper was presented rather than with a salutation. The eighteenth-century text does not contain an abstract. In terms of macro-structure, the

text is divided simply into paragraphs but not into sections, so there are no section headings either. But it is not only the format and style of the genre that have undergone substantial changes. Another explanation for the densification in the noun phrase has been sought in the process of text production. The advent of word-processors, in particular, has revolutionized writing. They allow more careful crafting and revision (they 'facilitate authors' abilities to manipulate' text) (Biber & Clark 2002: 63f.). It is not surprising, therefore, that we see similar structural changes in two genres that are subject to pressures to communicate efficiently in the written medium in the twentieth century: news and scientific writing. The changes that have affected the register of academic writing (from epistolary to research article, including the development of macrostructural elements such as the abstract, etc.) are so substantial that one might ask whether we are dealing with changes within a genre or to a different text type.

Authors' addresses:

Englisches Seminar

Universität Zürich

Plattenstrasse 47

CH-8032 Zürich

Switzerland

m.hundt@es.uzh.ch, gerold.schneider@es.uzh.ch

Linguistics and English Language

University of Manchester

Manchester M13 9PL

U.K.

david.denison@manchester.ac.uk

REFERENCES

Atkinson, Dwight. 1996. The 'Philosophical Transactions of the Royal Society of London', 1675-1975: A sociohistorical discourse analysis. *Language in Society* 25.3, 333-71.

Atkinson, Dwight. 1999. *Scientific discourse in sociohistorical context: 'The Philosophical Transactions of the Royal Society of London', 1675-1975*. London and Mahwah, NJ: Laurence Erlbaum.

Bain, Alexander. 1863. *An English grammar*. London: no publisher.

Ball, C. N. 1994. Automated text analysis: Cautionary tales. *Literary and Linguistic Computing* 9.4, 295-302.

Ball, Catherine N. 1996. A diachronic study of relative markers in spoken and written English. *Language Variation and Change* 8, 227-58.

Banks, David. 2008. *The development of scientific writing: Linguistic features and historical context* (Discussions in Functional Approaches to Language). London and Oakville CT: Equinox.

Barber, Charles. 1997. *Early Modern English*, 2nd edn. Edinburgh: Edinburgh University Press.

Biber, Douglas & Victoria Clark. 2002. Historical shifts in modification patterns with complex noun phrase structures. In Teresa Fanego, María José López-Couso & Javier Pérez-Guerra (eds.), *English historical syntax and morphology: Selected papers from 11 ICEHL, Santiago de Compostela, 7-11 September 2000* (Current Issues in Linguistic Theory 223), 43-66. Amsterdam and Philadelphia PA: John Benjamins.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style* (Cambridge Textbooks in Linguistics). Cambridge, etc.: Cambridge University Press.

Biber, Douglas & Bethany Gray. 2011. Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics* 15.2, 223-50.

Biber, Douglas, Jack Grieve & Gina Iberri-Shea. 2009. Noun phrase modification. In Günter Rohdenburg & Julia Schlüter (eds.), *One language, two grammars? Differences between British and American English* (Studies in English Language), 182-93. Cambridge: Cambridge University Press.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson.

Bisang, Walter. 2009. On the evolution of complexity - sometimes less is more in East and mainland Southeast Asia. In Geoffrey Sampson, David Gil & Peter Trudgill (eds.), *Language complexity as an evolving variable* (Oxford Studies in the Evolution of Language 13), 34-49. Oxford: Oxford University Press.

Cobbett, William. 1823. *A grammar of the English language, in a series of letters*. Oxford: Oxford University Press. Repr. in facsimile, 1984.

Crystal, David. 2003. *A dictionary of linguistics and phonetics*, 5th edn. Oxford and Malden MA: Blackwell.

Dekeyser, Xavier. 1984. Relativizers in Early Modern English: A dynamic quantitative study. In Jacek Fisiak (ed.), *Historical syntax*, 61-87. Paris and The Hague: Mouton.

Denison, David & Marianne Hundt. submitted. Defining relatives.

Fitzmaurice, Susan. 2000. *The Spectator*, the politics of social networks, and language standardisation in eighteenth-century England. In Laura Wright (ed.), *The development of Standard English 1300-1800*, 195-218. Cambridge: Cambridge University Press.

Fowler, H. W. 1926. *A dictionary of Modern English usage*. London: Oxford University Press.

Garner, Bryan A. 2003. *Garner's modern American usage*, 2nd edn. Oxford: Oxford University Press.

Geisler, Christer & Christine Johansson. 2002. Relativization in formal spoken American English. In Marko Modiano (ed.), *Studies in Mid-Atlantic English* (HS-institutionens skriftserie), 87-109. Gävle: Högskolan i Gävle.

Gilman, E. Ward (ed.) 1994. *Merriam-Webster's dictionary of English usage: The complete guide to problems of confused or disputed usage*. Springfield MA: Merriam-Webster.

Givón, Talmy & Masayoshi Shibatani (eds.). 2009. *Syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution* (Typological Studies in Language 85). Amsterdam: John Benjamins.

Gotti, Maurizio. 2003. *Specialized discourse: Linguistic features and changing conventions*. Bern: Peter Lang.

Grijzenhout, Janet. 1992. The change of relative *that* to *who* and *which* in late seventeenth-century comedies. *NOWELE* 20, 33-52.

Gut, Ulrike & Lilian Coronel. 2012. Relatives worldwide. In Marianne Hundt & Ulrike Gut (eds.), *Mapping unity and diversity world-wide: Corpus-based studies of new Englishes*, 215-41. Amsterdam and Philadelphia PA: John Benjamins.

Huddleston, Rodney, Geoffrey K. Pullum & Peter Peterson. 2002. Relative constructions and unbounded dependencies. In Rodney Huddleston & Geoffrey K. Pullum *et al.* (eds.), *The Cambridge grammar of the English language*, 1031-96. Cambridge: Cambridge University Press.

Hundt, Marianne. 2011. Relatives in scientific English: Variation across time and space. Paper presented at CLAVIER 11 conference: Tracking Language Change in Specialised and Professional Genres, Modena.

Hundt, Marianne, David Denison & Gerold Schneider. 2012. Retrieving relatives from historical data. *Literary and Linguistic Computing* 27.1, 3-16.

Hundt, Marianne & Geoffrey Leech. forthcoming,2012. Small is beautiful: On the value of standard reference corpora for observing recent grammatical change. In Terttu Nevalainen & Elizabeth Traugott (eds.), *The Oxford handbook of the history of English*. New York: Oxford University Press.

Johansson, Christine. 2006. Relativizers in nineteenth-century English. In Merja Kytö, Mats Rydén & Erik Smitterberg (eds.), *Nineteenth-century English: Stability and change* (Studies in English Language), 136-82. Cambridge: Cambridge University Press.

Leech, Geoffrey, Marianne Hundt, Christian Mair & Nicholas Smith. 2009. *Change in contemporary English: A grammatical study* (Studies in English Language). Cambridge: Cambridge University Press.

Lehmann, Christian. 1984. *Der Relativsatz* (Language Universals 3). Tübingen: Gunter Narr.

Montgomery, Michael. 1989. The standardization of English relative clauses. In Joseph B. jr Trahern (ed.), *Standardizing English: Essays in the history of language*

change: In honor of John Hurt Fisher (Tennessee Studies in Literature 31), 113-38. Knoxville TN: University of Tennessee Press.
Morris, Richard. 1895. *Historical outlines of English accidence*, 2nd edn. London: Macmillan.
Mustanoja, Tauno F. 1960. *A Middle English syntax*, vol. 1, *Parts of speech* (Mémoires de la Société Néophilologique de Helsinki 23). Helsinki: Société Néophilologique.
Nevalainen, Terttu. 2002. The rise of *who* in Early Modern English. In Patricia Poussa (ed.), *Relativization on the North Sea littoral*, 109-21. Munich: LINCOM Europa.
Pérez Guerra, Javier & Ana E. Martínez Insua. 2010a. Do some genres or text types become more complex than others? In Heidrun Dorgeloh & Anja Wanner (eds.), *Syntactic variation and genre*, 111-40. Berlin: Mouton de Gruyter.
Pérez Guerra, Javier & Ana E. Martínez Insua. 2010b. Enlarging noun phrases little by little: On structural complexity and modification in the history of English. In Aquilino Sánchez & Moisés Almela (eds.), *A mosaic of corpus linguistics: Selected approaches*, 193-210. Frankfurt-am-Main: Peter Lang.
Peters, Pam. 2004. *The Cambridge guide to English usage*. Cambridge, etc.: Cambridge University Press.
Rissanen, Matti. 1984. The choice of relative pronouns in 17th century American English. In Jacek Fisiak (ed.), *Historical syntax*, 417-35. Paris and The Hague: Mouton.
Romaine, Suzanne. 1980. The relative clause marker in Scots English: Diffusion, complexity, and style as dimensions of syntactic change. *Language in Society* 9, 221-47.
Rydén, Mats. 1984. När är en relativsats "nödvändig"? *Moderna Språk* 78, 19-22.
Schneider, Gerold. 2008. *Hybrid long-distance functional dependency parsing*. PhD dissertation, University of Zürich.
Sigley, Robert. 1997. *Choosing your relatives: Relative clauses in New Zealand English*. PhD dissertation, Victoria University.
Strunk, William, Jr & E. B. White. 1999. *The elements of style*, 4th edn. London and New York: Longman.
Taggart, Caroline & J. A. Wines. 2008. *My Grammar and I (or should that be 'Me'?)*. London: Michael O'Mara Books.
Tagliamonte, Sali. 2002. Variation and change in the British relative marker system. In Patricia Poussa (ed.), *Relativization on the North Sea littoral*, 147-65. Munich: LINCOM Europa.
Tottie, Gunnel. 1997a. Literacy and prescriptivism as determinants of linguistic change: A case study based on relativization strategies. In Uwe Böker & Hans Sauer (eds.), *Anglistentag 1996, Dresden: Proceedings*, 83-93. Trier: Wissenschaftlicher Verlag.
Tottie, Gunnel. 1997b. Overseas relatives: British-American differences in relative marker usage. In J. Aarts, Inge de Mönnink & H. Chr. Wekker (eds.), *Studies in English language research and teaching: In honor of Flor Aarts*, 153-65. Amsterdam and Atlanta GA: Rodopi.
Yáñez Bouza, Nuria. 2011. ARCHER past and present (1990-2010). *ICAME Journal* 35, 205-36.

APPENDIX

	NN sequence per nchunk	NNN sequence per nchunk	adj-adj sequence per nchunk
1700s	0.029489386	0.009753299	0.01113023
1800s	0.047641289	0.009925269	0.013078001
1900s	0.116795367	0.020511583	0.024975869

Table 1a: Complex premodifications per nchunk in ARCHER (BrE)

	NN sequence per nchunk	NNN sequence per nchunk	adj-adj sequence per nchunk
1700s	0.034411384	0.009573092	0.013971539
1800s	0.052977839	0.009926131	0.018351801
1900s	0.110417667	0.023715795	0.026788286

Table 1b: Complex premodifications per nchunk in ARCHER (AmE)